DOCUMENT RESUME

ED 129 884 TM 005 687

AUTHOR Durell, A. B.
TITLE A Computer Simulation Study of Measures for Validating Learning Hierarchies.
PUB DATE [Apr 74]
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (58th, Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Computer Assisted Instruction; *Computer Oriented Programs; *Educational Objectives; *Mathematical Models; Measurement Techniques; *Organization; Psychometrics; *Simulation; Statistical Analysis; *Validity
IDENTIFIERS *Learning Hierarchies

ABSTRACT

Learning hierarchies are networks of prerequisite relationships of instructional objectives. Seven measures of the validity of learning hierarchies were compared for their ability to identify correctly- and incorrectly-ordered hierarchies. A computer simulation model was used to generate stochastic data of known underlying structure. Analysis of variance processing of the data indicated that three of the measures provide stringent but useful tests of hierarchy validity. (Author/BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDPS are the best that can be made from the original. *
****

ED129884

A COMPUTER SIMULATION STUDY OF MEASURES FOR VALIDATING LEARNING HIERARCHIES

by

A.B. Durell[1]

The Ontario Institute for Studies in Education

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the Annual Meeting
of the
American Educational Research Association
Chicago, April 1974

TM005 687

1. Now at Department of Educational Psychology, Faculty of Education, University of Toronto.

Problem

Learning hierarchies (Gagné, 1965) are networks of prerequisite relationships of instructional objectives. Use of the term seems to have followed from the work of Gagné and Paradise (1961). Designers of systematic approaches to individualized instruction (Bolvin, 1968), diagnostic and achievement testing (Glaser & Nitko, 1971) and mastery learning (Bloom, 1971) often find the learning hierarchy concept to be useful. Computer-based instructional systems are often based on learning hierarchies. Examples may be found in computer-managed instruction (Sass, 1971), computer-based testing (Ferguson, 1970), and computer-assisted instruction (Hicks & Hunka, 1972).

The most useful approach to hierarchy generation is to begin with the terminal objective of an instructional sequence and ask the question "What would the learner have to be able to do in order to attain this objective?" as suggested by Gagné (1969). In this manner, behaviours prerequisite to performance of the terminal objective are identified. The question is repeated for each of the subordinate behaviours. Repetitive application of this heuristic generates a hierarchy of behaviours. The process is continued until it is reasonable to assume that the subordinate behaviours identified will be in the repetoires of all learners to be instructed.

Use of the foregoing procedure does not guarantee production of a hierarchy that is valid. That is, a network of objectives generated by logical analysis may prove to be pedegogically ineffective. Resnick (1973) distinguished between psychometric and transfer interpretations of hierarchy validity. Most attempts to construct learning hierarchies are grounded on a desire to identify prerequisite relationships which will provide transfer value between objectives in the hierarchy. However, most measures of hierarchy validity are based on psychometric evidence. Tests of transfer involve instructional intervention and the explicit comparison of the relative effectiveness of different orders of presenting instructional material. Psychometric measures use the performance patterns of learners for a hypotnesized hierarchy to test for the dependency relations which should exist if certain objectives truly are prerequisite to others. John Carroll, in the "Comments of Discussants" of the Resnick (1973) symposium, noted that psychometric evidence of hierarchical relationship is no guarantee that there is transfer value from one objective to another. However, psychometric measures do constitute a necessary, though not sufficient, condition for transfer to exist. Carroll suggested that psychometric indications of hierarchy validity are of value as hueristic devices in searching for hierarchies to test for transfer value. The transfer test of heirarchy validity

is an exacting one which takes a good deal of time. Therefore, psychometric measures of hierarchy validity are significant even though they are insufficient to certify the pedagogical worth of a hierarchy. All of the measures considered in this study are psychometric measures.

The purpose of this study is to consider the effectiveness of several psychometric measures of hierarchy validity in detecting correctly and incorrectly sequenced objectives. There is a practical difficulty in the way of performing such a test with any given hierarchy of instructional objectives. What is desired is to test the ability of various measures to indicate whether or not a hierarchy is valid. However, one cannot know if the hierarchy which is used to test the measures is valid. If it were possible to know if the hierarchy were valid, there would be no need for the measures. In this study a model of learning hierarchically related material was formulated and used as the basis of a computer program to generate data simulating that which might be produced by learners. The use of a model made it possible to specify the underlying structure of the data. In particular, it was possible to decree in advance that the hierarchy for which data would be generated was or was not valid.

Measures Considered

The prerequisite relationships which are assumed to exist in a hierarchy have suggested the use of scalogram

analysis (Guttman, 1944) and multiple scalogram analysis (Lingoes, 1963) as measures of hierarchy validity. One of the major difficulties in applying scaling techniques to hierarchy validation is that scaling techniques only indicate linear relationships while most hierarchies involve branches. Resnick and Wang (1969) found scaling procedures awkward to apply to a branched hierarchy.

Another class of measures, less mathematically sophisticated than scaling techniques, can be identified in the literature. This class of measures may be characterized as step-by-step measures as they use data concerning the mastery or non-mastery of objectives by learners to calculate numerical values associated with every transition from one hierarchy level to another. These measures do not produce any overall score of validity for an entire hierarchy, as scaling procedures do.

In the minimal hierarchy in Figure 1, mastery of objective B is assumed to be a prerequisite of mastery of objective A.

The possible patterns of results which a learner might produce for these two objectives are shown in Table 1. A + indicates mastery and a - indicates non-mastery. The pattern of results is represented by an ordered pair such as (-+). The pattern (-+) indicates non-mastery of the higher level objective (A) and mastery of the lower level objective (B).

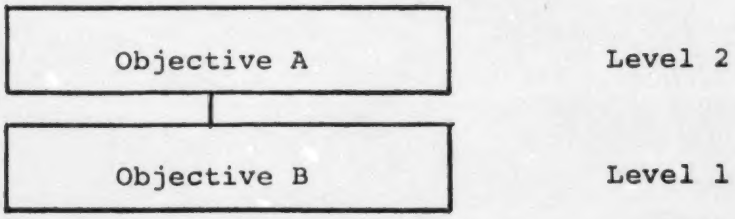

Figure 1: A two-objective hierarchy

TABLE 1

POSSIBLE RESPONSE PATTERNS FOR A LEARNER ON TWO OBJECTIVES

Lower Level objective (B)	Higher Level Objective (A): Mastery (+)	Higher Level Objective (A): Non-Mastery (-)
Mastery (+)	(+ +)	(- +)
Non-Mastery (-)	(+ -)	(- -)

In the calculational formulae to be presented for the various measures, the response patterns will be used to represent the number of learners demonstrating that particular pattern. The proportion of responses fitting, say, the (++) pattern will be symbolized by P(++). Thus,

$$P(++) = \frac{(++)}{(++) + (+-) + (-+) + (--)}$$

The six step-by-step measures selected for consideration in this study constitute all such measures suggested in the literature. The name of each measure, an abbreviation for future reference, its calculational formula, its originator, value range, and criterion value will be briefly stated.

1. Proportion of Positive Transfer (PPT). This measure was proposed by Gagné and Paradise (1961). The formula for PPT is

$$\frac{(++) + (--)}{(++) + (--) + (+-)}$$

PPT has a range of values from 0 to 1 and its criterion value is .90. That is, if the value of PPT calculated between any two objectives, for which a prerequisite relationship is hypothesized, is greater than or equal to .90, then PPT indicates that a hierarchical relationship exists.

2. Order Ratio (OR). Phillips (1971) devised the order ratio by adding the number of (-+) response patterns to the

numerator and denominator of PPT. Thus the formula for OR is

$$\frac{(++) + (--) + (-+)}{(++) + (--) + (-+) + (+-)}$$

OR has a range of values from) to 1 and a criterion value of .90. The apparent complexity of the formula for OR may obscure the fact that it is equal to

$$1 - P(+-).$$

3. Eisenberg-Walbesser Ratios. Three ratios, each of which tests for some desirable property of response patterns in a hierarchical relationship were proposed by Eisenberg and Walbesser (1971). All three ratios have a range of values from 0 to 1. It was suggested that a hierarchical relationship does not exist unless the values of all three ratios are greater than or equal to .85. For the purposes of this study, these three ratios were regarded as components of a single ratio called the Combined Eisenberg-Walbesser Ratio (CEW). CEW is equal to the minimum of the three component ratios and has a criterion value of .85.

The three component ratios and their formulae are:

Consistency Ratio

$$\frac{(++)}{(++) + (+-)}$$

Adequacy Ratio

$$\frac{(++)}{(++) + (-+)}$$

Completeness Ratio

$$\frac{(++)}{(++) + (--)}$$

4. Phi (PHI). Phillips (1971) used a phi coefficient as an indicator of hierarchical relationship. The phi coefficient is the product moment correlation coefficient for dichotomous data and its calculational formula, in terms of response patterns, is

$$\frac{(++)\,(--) - (-+)\,(+-)}{\sqrt{[(++) + (+-)]\,[(-+) + (--)]\,[(++) + (-+)]\,[(+-) + (--)]}}$$

PHI has a range of values from -1 to +1, and in this study a criterion value of .60.

5. Phi/Phimax (PPM). Resnick and Wang (1969) reported that Carroll was developing a validation procedure based on

$$\frac{\text{phi}}{\text{phimax}}$$

where phimax is the maximum value which phi could have, given the marginals of the contingency table. The calculation of PPM has been described by Cureton (1959). PPM has a range of values from -1 to +1 and a criterion value of .60 was used in this study

6. Difference Ratio (DR). This ratio was developed for this study. A complete description of the development of this ratio is given by Durel. (-973). The formula for DR is

$$\frac{(++)}{(++) + (-+) + (+-) + (--) + |(-+) - (--)|} - P(+-)$$

the range of values for DR is from -1 to +1 and the criterion value is .50.

7. Conditional Item Difficulty Index (CIDI). Airasian (1971) proposed a measure which differs from the other measures considered in this study. The value of CIDI for level n of a hierarchy is given by dividing the number of learners who have achieved mastery of all the objectives at levels 1 through n by the number of learners who have achieved mastery of all objectives at levels 1 through (n-1). The numbering of levels is from the bottom to the top of the hierarchy. For instance, objectives at level 5 of the hierarchy are prerequisites for objectives at level 6. CIDI has a range of values from 0 to 1 and a criterion value of .85 was used in this study.

Data Generation Model

A simplified model of learning hierarchically related material was developed. The model served as the basis for a computer program to generate simulated data for comparing the measures. The model made it possible to consider both the apparent state of a learner's mastery of a given

objective and the true underlying state. It is useful to distinguish between these conditions by referring to them as "indicated mastery" and "true mastery", respectively.

The model involves three parameters. The first parameter is called the "coefficient of transfer" (CT). CT is the probability of a learner having true mastery at a particular level of a hierarchy, given true mastery at all subordinate levels. In this study, CT was given the values .75, .85, and .95 to represent hierarchies demonstrating a range from weak to strong transfer. The other two parameters are probabilities of indicating mastery. PM is the probability that a learner will be judged to have mastery of an objective, given that he is in a state of true mastery of the objective. PN is the probability that a learner will be judged to have mastery of an objective, given that he is in a state of true non-mastery of the objective. PM was given the values .90 and .95 as instructional systems are usually designed to minimize false indications of non-mastery. Similarly, PN was given the values .05 and .10 as indications of false mastery are also minimized in most systems.

The model was used to generate indicated mastery states for each learner for each objective of a hierarchy. It was assumed that all learners had true mastery at the lowest level of a hierarchy. The indicated mastery would then be generated for the lowest level objective with a probability

of PM of indicating mastery. Then the true mastery state for the learner on the next objective of the hierarchy was generated with a probability of CT of having true mastery. For each objective for which a learner had true mastery, the indicated mastery state was generated with a probability of PM that mastery would be indicated.

When a state of non-mastery was generated for a particular objective of the hierarchy, the indicated mastery state was generated with a probability of PN that mastery would be indicated. Furthermore, once a learner entered a state of true non-mastery for a particular objective, he remained in a state of true non-mastery for all higher level objectives in strict adherence to the assumptions of learning hierarchy theory. Therefore, once a learner entered a state of true non-mastery it was no longer necessary to use the value of the parameter CT to generate the true mastery state for that learner for higher level objectives of the hierarchy.

The indicated mastery states for a learner were paired to give a response pattern for each transition from one hierarchy level to another. The response patterns were tallied over the whole set of simulated learners and used to calculate the values of the seven measures of learning hierarchy validity.

Method

The model was used to generate performance data on a stochastic basis for groups of 100 simulated learners each. The performance data were used to calculate values of each of the seven measures of hierarchy validity for each level transition in a hierarchy. An eleven-level hierarchy was used. Thus there would be ten level transitions in the hierarchy and so ten indications by each measure as to whether a hierarchical relationship existed between successive levels of the hierarchy. The use of the simulation made it possible to know the true underlying nature of the hierarchy. On that basis, each value of a validity measure could be evaluated as indicating a correct or incorrect decision concerning existence of a hierarchical relationship. Each measure indicated ten decisions. The number of correct decisions, which could range from 0 to 10, was the independent variable.

The parameters of the model were used as three factors in a factorial design. The probability values of the parameters were used as levels of the factors. The seven measures of hierarchy validity were used as a fourth factor with seven levels. Thus the experimental design was CTxPMxPNxMeasures which led to a 3x2x2x7 analysis of variance. Data were subjected to an arcsin transformation. Data for ten groups of 100 simulated subjects each were obtained for each of the 84 cells of the design.

Four experiments were performed involving a variety of arrangements of the objectives of the hierarchy to test the ability of the measures to detect errors in arrangements of objectives. In one experiment the hierarchy had the objectives in the correct order. Three experiments involved incorrect orderings of the objectives. In one, two objectives were out of correct order, in another three objectives were out of correct order, and in the third the objectives were ordered randomly.

Results

The analyses of variance carried out on the data for the four experiments produced significant effects for the Measures factor in every case. That is, there were significant differences in the abilities of the seven measures to make correct decisions concerning the presence of hierarchical relationships in the four different orderings of the objectives of a hierarchy. These results are summarized in Figure 2.

The mean numbers of correct decisions made showed that PPT and OR are generally less able to indicate correct decisions. In addition, the more incorrect the ordering of the objectives, the fewer correct decisions PPT and OR made. This would indicate that PPT and OR have a tendency to indicate the presence of hierarchical relationships which do not exist.

PPM performed slightly better than PPT and OR but not

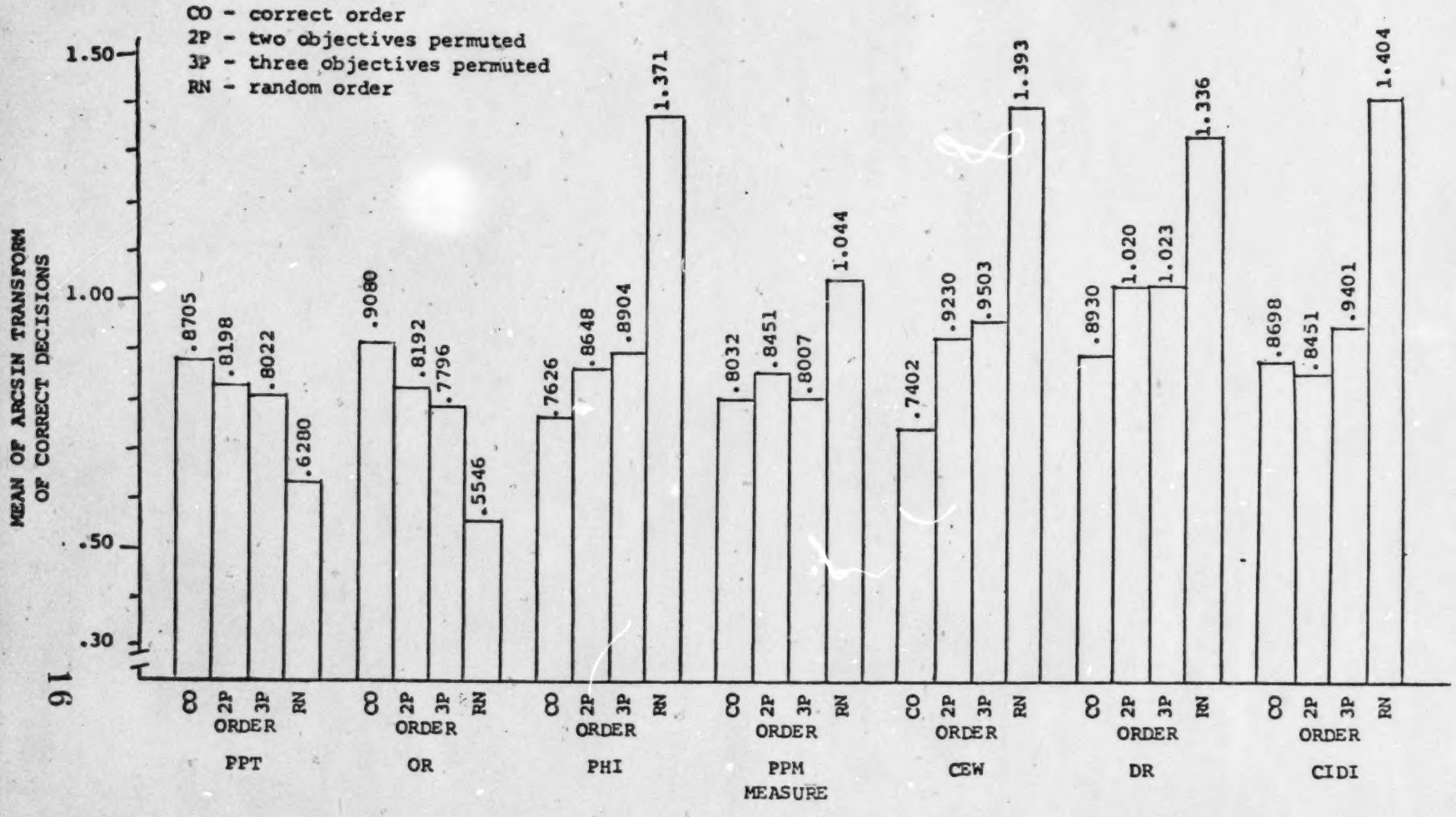

Figure 26: Means produced by measures for each hierarchy

as well as the other four measures. PHI produced moderately better results than PPM on the first three experiments and substantially better results on the fourth experiment.

CEW produced a moderate average improvement over PHI on the first three experiments and CIDI performed slightly better than CEW. DR produced the most consistent results overall.

The more incorrect the ordering of the objectives, the more correct decisions CEW and DR made. This would indicate that CEW and DR have a tendency to indicate a lack of hierarchical relationship even when such relationships might exist.

Further useful information was obtained by examining the interactions of the CT and Measures factors. PPT and OR tended to make correct decisions for the highest vale of the CT factor, but made many incorrect decisions for low values of the CT factor. Conversely, CEW and DR made relatively few correct decisions at the highest level of CT, but performed quite well at lower values of CT. These tendencies are of importance since CT is, in effect, an indication of the strength of the hierarchy. Consistent trends of this sort were not evident in the CTxMeasures interaction data over the four experiments for the other three measures (Durell, 1973).

Discussion

The tendency of PPT and OR to indicate a large number of correct decisions for a well-ordered hierarchy and for the high value of CT suggests that these measures are "liberal". That is, PPT and OR tend to produce high values for a wide range of frequencies of response patterns. This means that, for a hierarchy with incorrectly ordered objectives and/or low values of CT, PPT and OR will give incorrect indications of hierarchical relationship.

On the other hand, CEW and DR had a tendency to produce low values and therefore to indicate that hierarchical relationships did not exist. CEW and DR might be characterized as "conservative" measures of hierarchy validity. These two measures made many correct decisions for hierarchies with correctly ordered objectives and/or medium or low CT values. Thus it seems that CEW and DR are sufficient to detect many instances of lack of hierarchical relationship but do not perform as well in identifying instances in which hierarchical relationship does exist.

PHI, PPM, and CIDI did not have as clearly distinguishable characteristics as PPT, OR, CEW, and DR. PPM had a general poor ability to make correct decisions. PHI had some of the characteristics of CEW and DR, but was less effective than those two measures in making correct decisions. CIDI was intermediate between CEW and DR in overall ability to indicate correct decisions. CIDI was

less affected by variations in the value of CT than CEW and DR.

The general low ability of PPT, OR, and PPM to make correct decisions suggests that they are poor indicators of hierarchy validity. PHI performed better, but not as well as CEW, CIDI, and DR. In addition, values of PHI are somewhat difficult to compute.

CEW, DR, and CIDI demonstrated reasonable ability to make correct decisions. However, all of these three measures tended to be better at indicating lack of hierarchical relationship than presence of hierarchical relationship. Of all the faults which a measure of hierarchy validity may have, this tendency to be conservative is not a difficult one to deal with. At worst it means that a proposed hierarchy will be judged against a very stringent criterion. It is suspected that changing the criterion value for these measures might lead to an improvement in the ability of one or all of them to make correct decisions. The task of determining optimum criterion values may be approached through further simulation studies.

References

Airasian, P.W. A method for validating sequential instructional hierarchies. Educational Technology, December 1971, 11, 54-56.

Bloom, B.S. Mastery learning. In J.H. Block (Ed.) Mastery learning: Theory and practice. New York: Holt, Rinehart & Winston, 1971, 47-63.

Bolvin, J.O. Implications of the individualization of instruction for curriculum and instructional design. Audiovisual Instruction, March 1968, 13, 238-242.

Durell, A.B. A comparison of measures for validating learning hierarchies. Ph.D. Dissertation, University of Toronto, 1973.

Ferguson, R.L. A model for computer-assisted criterion-referenced measurement. Education, 1970, 9, 25-31.

Gagné, R.M. The conditions of learning. New York: Holt, Rinehart & Winston, 1965.

Gagné, R.M. Learning hierarchies. Educational Psychologist, 1968, 6, 1, 1-9.

Gagné, R.M. & Paradise, N.E. Abilities and learning sets in knowledge acquisition. Psychological monographs, 1961, 75 (Whole No. 518).

Glaser, R. & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.) Educational measurement. Second edition. Washington, D.C.: American Council on Education, 1971, 625-670.

Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-150.

Hicks, B.L. & Hunka, S. The teacher and the computer. Philadelphia: Saunders, 1972.

Lingoes, J.C. Multiple scalogram analysis: A set theoretic model for analyzing dichotomous items. Educational and Psychological Measurement, 1963, 23, 501-524.

Phillips, E.R. Validating learning hierarchies for sequencing mathematical tasks. Unpublished Doctoral Dissertation, Purdue University, 1971.

Resnick, L.B. (Ed.) Hierarchies in children's learning: A symposium. Instructional Science, 1973, 2, 311-362.

Resnick, L.B. & Wang, M.C. Approaches to the validation of learning hierarchies. In Proceedings of the Eighteenth Annual Western Regional Conference on Testing Problems, Educational Testing Service, 1969, 14-38.

Sass, R. E. A computer-based instructional management program for classroom use. Learning Research and Development Center, University of Pittsburgh, May 1971.